HOMEWORK 3 – GROUP 1

Problem 3.1

In this question, we are asked to prove that $A\boldsymbol{w} - b$ is a supergradient of the dual function of the primal function given in slide 11 of lecture 5. The group members' approach used here is correct. However, the final step of the proof (the fourth step of the green color text) is not clear. Technically speaking, to write that step, one should initially introduce that, for a given $\boldsymbol{\lambda}$, the minimum of $f(\boldsymbol{w}) + \boldsymbol{\lambda}^T(A\boldsymbol{w} - b)$ attaints at some $\boldsymbol{w}$. To avoid the notational confusion, we suggest that it would be better to initially state (instead of the second blue color equation of their solution) that the minimum of $f(\boldsymbol{w}) + \boldsymbol{\lambda}^T(A\boldsymbol{w} - b)$ attaints at $\boldsymbol{w} = \boldsymbol{w_0}$. Then, we can have that $g(\boldsymbol{\lambda}) = f(\boldsymbol{w_0}) + \boldsymbol{\lambda}^T(A\boldsymbol{w_0} - b)$. Hence, by using $\boldsymbol{w_0}$ instead of $\boldsymbol{w}$ in the second and third lines of their proof (green color text), they could yield the desired solution.

Problem 3.2

In this question, we need to analyze the convergence of dual ascent for L −smooth and μ −strongly convex f given in slide 11 of lecture 5. As the initial step, the group members have written the dual function in terms of the conjugate function f^* of f. However, the smoothness and strong convexity properties which they have stated here are not correct. Moreover, it is worth substantiating the related properties, other than directly stated. After stating the smoothness and strong convexity properties of f^*, they directly stated the rate of convergence of the dual ascent without analyzing the smoothness and strong concavity of the dual function. We suggest that they should analyze how they could derive the smoothness and strong convexity constants of the dual function because they are directly depending not only on μ and L but also on matrix A.

Problem 3.3

In this question, we are asked to use the dual decomposition method to solve the consensus optimization problem (P2) given in slide 21 of lecture 5 and compare it to the primal method (analytically or numerically) in terms of total communication cost and convergence rate on a random geometric communication graph. Here, group 1 has used a well-known method of solving this problem by representing the equality constraint using the square root of a gossip matrix. We suggest that it is better to cite the related literature for completeness. The related distributed dual decomposition algorithm is given by using an appropriate variable transformation. However, the rate of convergence is not explicitly analyzed analytically or numerically. We believe that a nice comparison with the primal method could be performed analytically. Moreover, we suggest that a numerical analysis also could be performed to get more understanding of the related algorithms with different step size rules and different communication graphs.